数据新闻发展与实践研究
——以澎湃“美数课”为例

郭 岩
（新华社 中国经济信息社有限公司，北京 100052）

摘 要：为研究国内数据新闻的发展与实践现状，本文选择以澎湃网《美数课》数据新闻栏目为研究对象，使用爬虫抓取了其自 2016 年以来发布的所有的数据新闻的标题、发布时间、关键词、点赞数和评论数，以及数据新闻的可视化表现形式，进行了详细的数据统计和分析，从内容聚焦、表现形式和传播效果等方面分析了该栏目在数据新闻方面的实践，并最后提出建议：通过注重选题和创新可视化表现形式等方面提升数据新闻的质量，发挥出数据新闻的特色和优势，以促进数据新闻的发展。

关键词：数据新闻；美数课；可视化表现形式；传播效果　　**中图分类号：**G210　　**文献标识码：**A
文章编号：1671-0134（2021）04-098-03　　**DOI：**10.19483/j.cnki.11-4653/n.2021.04.027
本文著录格式：郭岩 . 数据新闻发展与实践研究——以澎湃“美数课”为例 [J]. 中国传媒科技，2021（04）：98-100.

导语

数据新闻理念的产生，最早可追溯至 2006 年，并在 2010 年逐步兴起。在这一时期，数据新闻在业界逐渐得到了认可，并进入了政府和公众的视野，有“互联网之父”之称的蒂姆·伯纳斯－李（Tim Berners-Lee）宣称：分析数据将成为未来新闻的特征。[1]

本文选取用户数众多的澎湃网作为研究对象，爬取了其数据新闻栏目《美数课》近五年来发布的数据新闻数据，进行了详细的统计和分析，以探究数据新闻的特征、传播效果，以及其发展趋势和反思等，最后提出对于数据新闻发展和实践的建议。

1. 数据新闻研究背景

1.1　数据新闻概况

数据新闻是大数据时代新闻学发展形成的新领域，它以开放的数据为基础，将复杂、抽象、难懂的数据转化为形象、具体、生动的新闻报道。

在我国，数据新闻的兴起与国外基本保持了同样的步伐。2011 年，搜狐推出了《数字之道》栏目；次年，网易、新浪、腾讯先后创立了《数读》《图解天下》《数据控》；2013 年，财新网成立数据新闻中心，推出《数字说》栏目；2015 年，第一财经新媒体公司成立以财经数据新闻内容生产为主业的新媒体平台——DT 财经，澎湃网开设了《美数课》栏目。同时，我国主流媒体在数据新闻领域也进行了大胆尝试，人民网在 2013 年推出了由人民网要闻部出品的《图解新闻》频道，新华网在 2012 年启动数据新闻项目，推出了《数据新闻》栏目。[2]

1.2　《美数课》栏目实践现状

澎湃网的《美数课》栏目开设于 2015 年，其宗旨为“数字是骨骼，设计是灵魂。与新闻相关，又与新闻无关”。澎湃网将该栏目归类于“时事”类别，发布的数据新闻主题更多的是关于当前社会热点话题，或者针对突发事件的报道分析。栏目每个月平均发布 14 篇数据新闻，新闻的可视化表现形式包括静态信息图、交互可视化、视频和动画等。读者可以对发布的数据新闻进行点赞、评论、分享和转发操作。

此次对数据新闻发展与实践的研究选择以澎湃网的《美数课》栏目为研究对象，一是因为其发布数据新闻的频率较高，二是因为其用户数较多且比较活跃。综合考虑其数据新闻的制作、运营能力和传播效果，在业界属于较高的水平，可以提供较为全面地分析素材。[3]

1.3　数据选取说明

本次数据抓取时间为 2021 年 3 月 1 日，抓取了《美数课》栏目所有数据新闻的标题、发布时间、关键词、点赞数和评论数，以及数据新闻的可视化表现形式。筛除部分邀请读者投票的内容，共有 795 篇数据新闻，可抓取到的首篇数据新闻发布时间为 2016 年 6 月。

2. 可视化内容生产分析

2.1　内容聚焦热点

按照数据新闻的选题性质划分，数据新闻可划分为事件选题型数据新闻与话题选题型数据新闻。事件选题型数据新闻是指将报道选题聚焦于某一新闻事件，对事件呈现出来的具有新闻价值的数据进行解读和呈现，这种新闻最常出现在重大新闻事件的报道中，例如会议、赛事、灾害性事件等。话题选题型数据新闻是指将报道聚焦于某类新闻话题，围绕此话题收集数据，并对之进行解读和呈现的数据报道。

经对抓取到的 795 篇数据新闻统计，绝大部分新闻皆为话题选题型，事件选题型较少，占比为 7%。这是因为重大的新闻事件的发生具有不确定，且出现的频率较低。因此，数据新闻的制作往往选取比较热门的话题和具有探讨性的主题，以期引发读者的阅读兴趣。

同时，在对抓取到的数据新闻的标题和关键词进行词频处理和分析后，发现出现在前十的高频词分别是：“新冠”“疫情”“美国”“奥运”“大选”“收入”“电影”“事故”“游戏”和“高考”。可见，《美数课》栏目的选题聚焦于社会热点，相较于传统文字新闻，数据新闻通过采集、挖掘相关数据，从数据的角度对社会热点话题进行了更进一步地分析报道，站位更加客观，可引发读者进一步思考，帮助读者对相关问题理解得更加透彻。

2.2 可视化表现形式

数据新闻的可视化表现形式可以包括：静态信息图、交互可视化、视频 / 动画、VR/AR 新闻等等。本文抓取的数据样本中，其表现形式主要使用了静态信息图、视频 / 动画和 H5 交互页面这三种形式。其中，H5 交互页面和视频 / 动画这两种较少，占比分别是 6% 和 9%。而静态信息图占了绝大多数，百分比是 85%。

交互可视化和视频类的表现形式，更加活泼、新颖，互动效果更好，增加了与读者的交互，更容易引发读者的兴趣，但是相对的制作更为困难，制作周期更长。因此，《美数课》更多的是选择了静态信息图的表现形式，间以 h5 交互和视频 / 动画可视化。这样，在大胆尝试使用交互可视化和视频 / 动画可视化表现形式的同时，也通过使用简单快捷的静态信息图可视化表现形式以保证数据新闻的时效性和较高的更新频率。

3. 传播效果分析

数据新闻的传播效果，可以通过新闻的点赞数和评论数来衡量。本文抓取了《美数课》栏目每一篇数据新闻的点赞数和评论数。由于 H5 交互页面没有点赞和评论的功能，因此对传播效果的分析，从 795 篇样本数据中剔除 48 篇 H5 交互的数据新闻，剩余 747 篇。

由点赞到评论，是读者对数据新闻由浅入深的参与形式。在完成阅读后，首先读者对读到的内容比较满意，就会选择点赞。更进一步的，如果对读到的内容有共鸣，或者产生了更深层次地理解和自己的见解，那读者就会去评论这篇新闻，或者去点评其他读者的评论内容。这是一种自然产生的与新闻交互的逐渐深入。

3.1 点赞统计

抓取的样本数据中，每篇数据新闻的平均点赞数是 190 个赞，点赞中位数是 89 个赞。其中，点赞数最多的新闻获得了 3587 个赞，点赞数最少的新闻仅获得了 1 个赞。点赞数排名前十的数据新闻如下表 1 所示。

表 1 点赞数前十的数据新闻

序号	标题	发布时间	点赞数
1	763 例确诊患者的故事，还原新冠病毒向全国扩散的路径	2020/2/5	3587
2	一张思维导图看懂民法典，它能保障你我的哪些权益？	2020/5/27	2259
3	数说马拉松“破 2”：他跑这 42 公里的速度，可能比你骑车快	2019/10/13	2129
4	AR ｜ 70 年来中国的城镇化进程	2019/10/6	1792
5	全国出院率过半 ｜ 一个月来，这张治愈地图在逐渐变绿	2020/3/1	1467
6	武汉解封倒计时，要走出心理的阴霾还需多久？	2020/4/7	1427
7	全球华人“店铺关门有家难回”？来看假消息是如何批量生产的	2020/3/18	1323
8	“钻石公主”号上新冠肺炎感染比例已高达 14.6%	2020/2/18	1282
9	无症状感染者纳入疫情通报，我们对该群体的认知到哪一步了？	2020/4/2	1197
10	女排豪取 10 连胜世界杯卫冕成功，数说 70 年女排辉煌战绩	2019/9/29	1146

在这 10 篇数据新闻中，第 1、5、6、7、8、9 篇，共 6 篇新闻都是关于新冠肺炎疫情的话题，可见疫情期间，读者对于疫情的数据新闻比较关注，表明社会热点话题更能引起读者的兴趣，传播效果更佳。另外，点赞数前十的新闻全部是 2019 年和 2020 年发布的，且 2020 年发布的有 7 篇。这表明，澎湃网的数据新闻栏目，近几年逐渐得到了读者的关注和认可，积累的用户数呈增加的趋势。

将点赞数按照数据新闻的发布时间画出趋势图，也可看出数据新闻的点赞数有逐年升高的趋势，也说明数据新闻的读者逐年增多。详情如下图 1 所示。

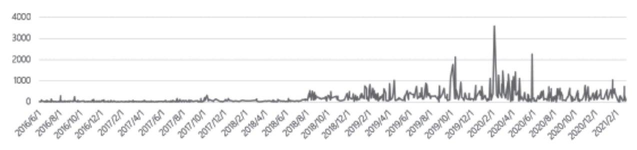


图 1 点赞数时间趋势

3.2 评论统计

抓取的样本数据中，每篇数据新闻的平均评论数是 99 个，评论中位数是 31 个。其中，评论数最多的新闻获得了 4590 个评论，有 26 篇数据新闻没有获得评论。评论数排名前十的数据新闻如下表 2 所示。

表 2 评论数前十的数据新闻

序号	标题	发布时间	评论数
1	南方阴雨大赛 ｜ 60 年的数据告诉你：这是太阳流浪最久的一次	2019/3/5	4590
2	一图看懂寿光洪灾，上游泄洪导致下游村庄被淹	2018/8/23	2593
3	一图看懂 ｜ 湖北等 31 省份启动的一级响应是什么？	2020/1/24	2110
4	数据起底无限极：何以成为中国最大直销公司	2019/1/19	1613
5	53027 条留言背后，网络树洞里绝望者的自救与互助	2019/9/10	1554
6	涨知识 ｜ 考了那么多年四六级，你知道自己比别人多交多少钱？	2016/10/11	1306
7	数说 ｜ 徐明朝拒向 Yamy 道歉，多少人遇到过职场 PUA ?	2020/7/22	1218
8	图解江歌案“罗生门”：控辩双方描述的“案情”有哪些分歧？	2017/12/11	1136
9	算一算，不再被忽视的家务劳动到底值多少钱？	2021/2/23	1095
10	数说高温② ｜ 65 年的数据告诉你：新十大火炉是如何炼成的？	2017/7/21	963

在这 10 篇数据新闻中，第 1、2 和 10 篇，是关于洪水和高温等灾害性天气的，第 3 篇是关于新冠肺炎疫情的话题，其他几篇也大多数是关于负面新闻的话题，可见社会负面的话题更能引发读者的讨论，更容易成为热点、爆点。关注产生力量，围观改变世界，读者对负面新闻的关注和讨论，往往有助于让事件朝好的方向发展，甚至推动社会进步，从而使负面新闻发挥出正面影响，这也体现着数据新闻发挥的价值。另外，评论数前十的

新闻，发布于2021、2020年、2019年、2018年、2017年和2016年的分别有1篇、2篇、3篇、1篇、2篇和1篇，没有明显的时间规律。

将评论数按照数据新闻的发布时间画出趋势图，也可看出数据新闻的评论数相较于点赞数没有明显的时间趋势。详情如下图2所示。

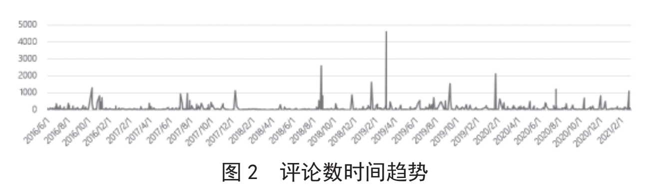


图2 评论数时间趋势

3.3 点赞与评论的传播效果分析

为了更客观地分析点赞数和评论数的相关程度，本文使用皮尔逊系数计算了两者的相关性。在统计学中，皮尔逊相关系数（Pearson correlation coefficient），[4]是用于度量两个变量X和Y之间的相关（线性相关），其值介于–1与1之间。两个变量之间的皮尔逊相关系数定义为两个变量之间的协方差和标准差的商，公式如下：

$$\rho_{X,Y}=\frac{cov(X,Y)}{\sigma_X\sigma_Y}=\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$$

相关系数的绝对值越大，相关性越强，相关系数越接近于1或–1，相关度越强，相关系数越接近于0，相关度越弱。通常情况下通过以下取值范围判断变量的相关强度：0.8–1.0极强相关；0.6–0.8强相关；0.4–0.6中等程度相关；0.2–0.4弱相关；0.0–0.2极弱相关或无相关。

经计算，点赞数和评论数的皮尔逊系数为0.2，属于弱相关。首先，点赞和评论都代表了读者对新闻的认可程度和感兴趣程度，因此，二者有一定的相关程度，且是正相关。点赞数越多，相对的评论数也会越多。但是点赞数和评论数两个变量的相关程度并不算高，且二者的时间趋势图也并不吻合，这是因为每个读者只可以点赞一次，却可以反复评论、评论多次，并可对别的读者的点评进行评论。近几年读者的增多，带来点赞数的增加，因此点赞数有逐年升高的趋势。而评论数则更多地反映出文章的热度。当某篇数据新闻发布出来之后，如果成为热点，其获得的评论数会迅速超过点赞数，传播效果成倍增加。因而，评论数的时间趋势图中，其峰值分散出现。

4. 数据新闻实践与发展的思考

通过前面对澎湃网《美数课》栏目的分析，可以发现，即使在前期用户数积累的阶段，如果做出了好的数据新闻，依然会产生评论数的峰值，传播效果明显。这表明，在注重数据新闻的更新量的基础上，更应注重数据新闻的质。通过制作更多的热点数据新闻，可以带来更多的点赞数、评论数，以及分享数和转发数，引发传播效果的指数增加，从而可以迅速提升栏目的知名度。

本文认为，提升数据新闻的质量，可以从以下几个方面着手。

4.1 选题迎合读者偏好

好的选题是数据新闻制作的第一步。数据新闻的选题，首先应顺应读者的偏好，引发读者阅读的兴趣。同时，数据新闻也有其鲜明的数据特征，是基于挖掘、分析数据为基础，有传统新闻无法替代的客观性、逻辑性，可以从更新颖的数据角度，对新闻事件、话题进行分析和阐述。因此，数据新闻的选题，应结合社会热点话题或者新闻事件，并可结合用户的点赞、评论数对用户的偏好进行分析，从中甄选出适合使用数据阐释的选题，进行数据新闻的生产。

4.2 创新可视化表现形式

澎湃网的《美数课》栏目，多数采用的是简单快捷的静态信息图，采用的其他比如视频、交互式表现形式等较少。新颖、丰富的多维数据新闻表现形式，可以有效提升用户体验，并且也是数据新闻相较于传统文字报道新闻的优势所在。考虑到非静态的表现形式相对需花费更多的制作时间，因此，建议综合考虑数据新闻的时效性与表现形式的多样性，并关注提升媒体的技术力量，提高数据新闻制作效率，更充分地展现出数据新闻的特色与优势。

结语

数据新闻是新闻与技术的结合。在进入大数据时代后，大量的数据增加了人们从中获取有价值信息的难度。数据新闻的产生，帮助读者挖掘海量数据中隐藏的信息，将有价值的部分筛选、梳理出来，并使用多样化的表现形式，展示给读者。数据的客观性，可以使读者对新闻事件有更深刻地理解，这是传统文字报道所无法传达的。国内许多媒体已经在数据新闻的制作方面积累了多年的实践经验。本文通过对澎湃网的《美数课》栏目进行数据抓取，对其在数据新闻实践方面进行了分析。结合分析的结果，本文建议数据新闻的制作应注重选题，并通过提升技术力量，尝试创新多维的可视化表现形式，促进数据新闻的进一步发展，以更好的服务广大读者。[5] 媒

参考文献

[1] 方洁.数据新闻概论[M].北京：中国人民大学出版社，2019，2.

[2] 王琼，徐园等.中国数据新闻发展报告（2018~2019）[M].北京：社会科学文献出版社，2020，1.

[3] 战贵桦.数据新闻报道疫情的实践探析——以财新网“数字说”为例[J].中国传媒科技，2020(9):99-101.

[4] 彭海.皮尔逊相关系数应用于医学信号相关度测量[J].电子世界，2017(7):163.

[5] 张超.释放数据的力量－数据新闻生产与伦理研究[M].北京：中国人民大学出版社，2020，1.

作者简介：郭岩（1988-），女，山东，工程师，研究方向：数据分析与管理。

（责任编辑：张晓婧）